

11459ROUS01U

METHOD AND APPARATUS FOR TRANSMITTING REAL-TIME DATA IN MULTI-
ACCESS SYSTEMS

FIELD OF THE INVENTION

5 This invention relates generally to the transmission of data in multi-access systems and more particularly to a method and apparatus for transmitting real-time data in multi-access systems.

10 BACKGROUND ART

Over the past few years, various multi-access systems have been developed in response to user demands for systems that can offer ready access to a wide variety of real-time or delay critical packet switched network services. Examples of these 15 services include voice over Internet protocol (VoIP), cable TV or telephony services.

In the vast majority of conventional multi-access systems which provide these types of services, transmission resources are typically assigned to users during periods they 20 actually have information to transmit. For example, in time assigned speech interpolation (TASI) systems where at any given time, multiple users engaged in different audio conversations share a limited number of transmission channels, channels are only allocated to each user during active speech segments or 25 bursts.

When a user initiates a speech segment in these systems, the speech segment is received at a statistical multiplexor which proceeds to allocate channels to transmit the speech segment. When the user enters periods of silence or 30 inactivity, the channels allocated are substantially reduced and typically re-allocated to other users or provisioned for control transmissions. This dynamic allocation of the available transmission resources also known as statistical multiplexing is

11459ROU\$01U

commonly used in multi-access systems to increase traffic capacity and more importantly, to maximize the use of transmission resources which are often limited.

When a user initiates a new speech segment and 5 switches from a state of inactivity to a state of activity, there is usually some delay before the necessary transmission resources can be allocated. This delay may result in situations where at the beginning of each speech segment, information is sent to a concentrator or multiplexor and is ready to be 10 transmitted but the channel resources necessary for its transmission are not yet available. In conventional systems, the information ready to be transmitted before channel resources become available is typically discarded.

However, because information is discarded, speech 15 segments are clipped at the onset causing information contained therein to be lost. In some systems, it has been shown that segments can be clipped for up to 40 milliseconds. Such clipping can severely disrupt user conversations, particularly where frequent pauses and silence periods occur.

20 This problem can also arise in multi-access wireless systems providing the same or other types of real-time services. In a multi-access wireless system providing audio services for example, delays in obtaining the appropriate radio resources are inevitable. Because of these delays, video segments or bursts 25 may be clipped. Again, this clipping may as a result substantially damage or distort entire transmissions.

In order to avoid clipping, some multi-access systems delay transmission until channel resources become available. Unfortunately, adding delays to avoid clipping may be 30 inappropriate. For example, adding delays during an audio conversation affect the entire dynamic of the conversation. In wireless systems, these delays considerably disrupt voice

11459ROUS01U

transmissions and reduce quality, sometimes below what is considered acceptable.

Therefore, when allocating radio resources in multi-access systems for the transmission of real-time or delay critical data such as, for example, audio or video information, it would be desirable to reduce delays and eliminate clipping to prevent transmission disruptions.

SUMMARY OF THE INVENTION

10 The present invention addresses these issues and to this end provides a methodology and apparatus to mitigate one or more of the present limitations in this art.

The invention provides a method and apparatus for transmitting real-time data in a multi-access system which substantially eliminates onset clipping of the data transmitted while reducing transmission delays. Generally, the invention can be incorporated in any multi-access system where transmission resources are allocated only when there is information to transmit.

20 According to a broad aspect, the invention provides a method of transmitting which includes detecting the start of an information segment being generated in real-time, editing and buffering the information segment or a first representation thereof to produce a second representation and, after 25 transmission resources have been allocated, starting to transmit the second representation whereby the editing and buffering is done to compensate for transmission resource allocation delays.

With respect to this particular aspect, the editing and buffering of the information segment can be performed with 30 or without other processing steps in different sequences including in particular editing first and then buffering or alternatively buffering first and then editing. Further, the editing and buffering can each be done on different

11459ROUS01U

representations of the information segment including the information segment as detected or as subsequently coded in frames.

The editing can be performed in a variety of ways including time compressing the segment or removing redundant frames therefrom if the segment is coded in frames first prior to any editing. According to the invention, time compressing the information segment preferably consists of removing repetitions and/or short pauses present in the segment. On the other hand, if the editing consists of removing redundant frames, the redundant frames removed are preferably those which contain repetitions and/or short pauses.

According to another broad aspect, the invention provides an apparatus to transmit information which includes an information detector operable to detect incoming information segments to transmit, an information editor operable to edit each information segment detected so as to produce a respective shortened information segment, a buffer operable to buffer each shortened information segment until transmission resources are allocated to produce a buffered information segment, and a transmitter operable to transmit each buffered information segment.

According to a preferred embodiment, the invention is incorporated in a multi-access wireless system for the upstream transmission of voice from a mobile station to a base station. According to the preferred embodiment, speech data received at the mobile station is edited to discard perceptually insignificant portions of the speech segment as it is received. The edited speech is then buffered to await transmission while the media access control (MAC) protocol layer acquires an allocation of transmission resources. By editing and buffering speech data as it is received, clipping of speech segments can be eliminated while reducing transmission delays.

11459ROUS01U

A variety of techniques may be used to edit and buffer the speech data as it is received to prevent onset clipping and reduce transmission delays. In the preferred embodiment, the speech data received is time-compressed with a speech/pause editor to remove repetitive segments and shorten pauses. The time-compressed speech is then coded in frames and the frames are placed in a buffer to await transmission. According to another preferred embodiment, the speech data received is coded in frames first. A speech frame editor examines the speech frames to discard frames deemed redundant. The frames which are not discarded by the speech frame editor are then placed in a buffer until ready to be transmitted.

Advantageously, by buffering the speech data received until the necessary transmission resources have been allocated, no meaningful speech information is lost and speech segments can be transmitted without any onset clipping. Another advantage of the invention is that by initially editing out perceptually insignificant portions of the speech data as it is received, the segments can be transmitted in a shorter time period to compensate for speech detection and resource allocation delays and reduce transmission delays.

The invention can advantageously be used for a variety of voice services such as for example, Enhanced Data for Global Evolution, voice over Internet Protocol (VoIP) services and audio conferencing. In addition, the invention can also be used for many other real-time services such as video conferencing.

BRIEF DESCRIPTION OF THE DRAWINGS

FIG. 1 is a block diagram of a typical multi-access wireless system;

FIG. 2 is a block diagram of a mobile station of the multi-access wireless system of figure 1;

FIG. 3 is a block diagram of the digital signal

11459ROUS01U

processor (DSP) block of Figure 2 showing in particular a speech encoder which edits and buffers speech data according to a preferred embodiment of the invention;

FIG. 4 is a timing diagram showing the editing and buffering of a speech segment by the speech encoder of Figure 3;

FIG. 5 is another block diagram of the DSP block of Figure 2 showing in particular a speech encoder which edits and buffers speech data according to another preferred embodiment of the invention;

10 FIG. 6 is a timing diagram showing the editing and buffering of a speech segment by the speech encoder of Figure 5; and

FIG. 7 is a block diagram of a protocol stack used for the transmission of speech data in the multi-access wireless 15 system of Figure 1.

DETAILED DESCRIPTION OF THE PREFERRED EMBODIMENTS

Embodiments of the invention provide a method and apparatus for transmitting real-time data in a multi-access 20 system which eliminates onset clipping of the data transmitted while reducing transmission delays. The invention can be incorporated in any multi-access system in which transmission resources are allocated only when there is information to transmit. For example, the invention can be incorporated in 25 multi-access wireless systems where radio resources used for transmission are allocated to users only during active speech periods.

An example of a multi-access wireless system which allocates radio resources only during active speech segments is 30 illustrated in Figure 1 as generally indicated by 10. In the wireless system 10 shown, radio coverage is divided into cells 12, 14, 16, 18, 20, 22 and 24 (only seven shown) where each cell 12, 14, 16, 18, 20, 22, 24 is assigned a number of available

11459ROUS01U

radio frequency (RF) transmission resources. These resources can be for example, fundamental or supplemental channels in code division multiple access (CDMA) or time slots in time division multiple access (TDMA). Generally, the nature of these 5 resources will depend on the type of wireless RF modulation technology employed. For clarity and generality, the resources available for transmission in each cell 12, 14, 16, 18, 20, 22, 24 of the network 10 are hereinafter referred to as transmission resources or simply resources.

10 Each cell 12, 14, 16, 18, 20, 22, 24 is serviced by a respective base station 32, 34, 36, 38, 40, 42, 44 which in turn is controlled by a mobile switching centre (MSC) 46. The MSC 46 provides external connectivity to other networks and systems such as the Internet or a public switched telephone network 15 (PSTN). Mobile stations 26, 28 (only two shown) communicate over wireless connections with the base station or base stations of the cells in which the mobile stations are located, base stations 34 and 42 in the illustrated example. With these connections, users of the network 10 can have access to standard 20 telephony services or other audio services such as Enhanced Data for Global Evolution or voice over Internet Protocol (VoIP) services.

In each cell 12, 14, 16, 18, 20, 22, 24 of the network 10, upstream and downstream communications are co-ordinated by 25 the resident base station 32, 34, 36, 38, 40, 42, 44. Each base station 32, 34, 36, 38, 40, 42, 44 controls access to the transmission resources available within their respective cell 12, 14, 16, 18, 20, 22, 24. With this control, the base stations 32, 34, 36, 38, 40, 42, 44 can manage their own 30 downstream transmissions and allocate resources to mobile stations 26, 28 within their cells 12, 14, 16, 18, 20, 22, 24 for the transmission of upstream data with the assistance of the MSC 46 as appropriate.

11459ROUS01U

Generally, in order to transmit upstream data, the mobile stations 26, 28 must request an allocation of transmission resources available within their respective cell 12, 14, 16, 18, 20, 22, 24. Considering upstream communications 5 in cell 14 for example, the mobile station 26 will generally not send any upstream data to the base station 34 unless the base station 34 allocates transmission resources to the mobile station 26 first. To maximize use of the transmission resources available in cell 14, resources are allocated to the mobile 10 station 26 on an information basis i.e. only when the mobile station 26 has active information to transmit or during active speech periods.

When a user initiates a speech segment at the beginning of a call, the mobile station 26 must obtain a 15 resource allocation from the base station 34 before it can begin to transmit the segment. However, the mobile station 26 does not retain the allocation for the entire duration of the call. During periods of inactivity or silence, the base station 34 re-allocates the resources to other users within the cell 14 or for 20 other purposes. This dynamic allocation also known as statistical multiplexing is used for upstream communications in the wireless network 10 to increase traffic capacity and maximize the use of the transmission resources available for upstream data in the cell 14.

25 The following will now describe the invention in more detail by way of example only in relation to the upstream transmission of voice from the mobile station 26 to the base station 34. It is to be understood that this description also applies upstream transmissions between other mobile stations and 30 base stations present in the network 10.

Figure 2 illustrates the mobile station 26 in further detail. As is conventional, the mobile station 26 includes multiple elements and devices which perform various functions

11459RQUS01U

including upstream communications, downstream communications, control, user interfacing and call set-up. For upstream communications, the mobile station 26 has a microphone 50, an analog-to-digital converter (ADC) 52, a speech encoder 55 implemented in a digital signal processor (DSP) unit 54 with a modulator 59, a transmitter 56, a power amplifier 58, a duplexer 60 and a radio antenna 62 all interconnected in series. For downstream communications, the mobile station 26 uses the antenna 62 and the duplexer 60 interconnected in series with a receiver 64, a demodulator 65 implemented in the DSP unit 54 with a speech decoder 57, a digital-to-analog converter 66 and a speaker 68. Transmission control is provided by a micro-controller unit 70 which is connected to the speech encoder 55 and the speech decoder 57 of the DSP unit 54.

According to the preferred embodiment, the speech encoder 55, the modulator 59, the speech decoder 57 and the demodulator 65 may be implemented other than within the DSP unit 54. For example, these components could alternatively be implemented with customized hardware without the need for a DSP unit 54.

In addition, the mobile station 26 also has a number of standard user interface and call set-up elements all controlled by the micro-controller unit 70 which include a human machine interface 72, a keypad 74 and a display unit 76 all interconnected in a standard fashion. These devices are well known in the art and are not described here in any detail.

When a user produces a speech segment at the mobile station 26 during an audio call, the speech is captured in the microphone 50 and digitized by the ADC 52 at a sufficiently high rate (e.g. 8000 8-bit samples per seconds) to digitize the entire user speech spectrum. Upon detecting the digitized speech, the speech encoder 55 of the DSP unit 54 codes the digitized speech data detected and places the coded speech in

11459ROUS01U

packets for transmission (further details below). Typically, in so doing, the DSP unit 54 reduces the bit rate to a much lower rate for radio transmission.

At the same time, the speech encoder 55 sends a message to the base station 34 (through the transmitter 56, the power amplifier 58, the duplexer 60 and the antenna 62) requesting transmission resources to transmit the speech segment being received. In response, the base station 34 allocates sufficient resources to the mobile station 26. Preferably, 10 these messages between the mobile station 26 and the base station 34 are exchanged using the media access control (MAC) protocol. In multi-access wireless networks such as the network 10, the MAC protocol is commonly used to co-ordinate and multiplex access to shared transmission resources by multiple 15 users.

When a response allocating transmission resources to the mobile station 26 is received (hereinafter also referred to as the MAC response or the MAC access), the speech encoder 55 begins to process the speech information detected in a manner to 20 substantially eliminate speech clipping (further details below). The speech encoder 55 forwards the processed speech information in the form of packets to the modulator 56 for modulation. In the modulator 56, the packetized data is modulated and then passed to the transmitter 56 for transmission to the base 25 station 34. When the user enters a period of silence or inactivity following the speech segment, the resources allocated for the segment are substantially reduced and re-allocated to other users within the cell 14.

In the wireless system 10 as in most other wireless systems, the resources allocated for transmitting a particular speech segment are not released immediately after the end of the segment. Typically, the resources are "held" for a certain time

11459ROUS01U

after the transmission of a speech segment. This holding time can be as long as a few seconds.

When the user initiates a new speech segment and switches from a state of inactivity to a state of activity, 5 there is usually some delay before the speech encoder 55 can detect the speech and proceed to request transmission resources for its transmission. More importantly, when the speech encoder 55 sends an allocation request upon detecting the beginning of a 10 speech segment, there can be a substantial delay before the speech encoder 55 can acquire MAC access and begin transmission of the segment. These delays may result in situations where at the beginning of each speech segment, speech frames are ready to transmit but the resources necessary for their transmission are not yet available.

15 In conventional wireless systems, the frames ready to transmit before MAC access are typically discarded. However, because these frames are discarded, speech segments are clipped at the onset causing information contained therein to be lost. In some conventional systems, it has been shown that segments 20 can be clipped for up to 40 milliseconds. Such clipping can severely disrupt user conversations, particularly where frequent pauses and silence periods occur.

According to the invention, when a speech segment is detected, the speech encoder 55 proceeds to edit the digitized 25 speech data as it is received from the ADC 52 to remove perceptually insignificant portions. The speech encoder 55 then places the edited speech in a buffer to await transmission until the proper transmission resources have been allocated and the edited speech data can be transmitted. In contrast to 30 conventional transmission methods which clip segments at the onset and can cause as a result important information contained therein to be lost, the present invention removes perceptually insignificant speech portions instead to catch up on the delays

11459ROUS01U

incurred in transmitting the information which would otherwise be clipped. According to the invention, the editing is deactivated only when sufficient time savings have been achieved to compensate for the additional time required to buffer and transmit the information which would have otherwise been discarded. As will be explained below in further detail, by editing and buffering the digitized speech data in the speech encoder 55, clipping of speech segments can be eliminated while reducing segment transmission delays. Once

10 A variety of techniques may be used to discard perceptually insignificant speech portions and buffer the edited speech data to prevent clipping while reducing segment transmission delays. According to a preferred embodiment, the speech data received is time-compressed with a speech/pause editor to remove repetitive portions and shorten pauses. Then, 15 the time-compressed speech is coded in frames with a speech coder and placed in a buffer to await transmission.

Figure 3 shows in more detail the DSP unit 54 of Figure 2 including in particular a preferred embodiment for the speech encoder 55 which can be used to time compress and buffer digitized speech data as it is received from the ADC 52. In this particular embodiment, the speech encoder 55 has an optional noise reduction unit 100 connected to receive the output of the ADC 52. The noise reduction unit 100 is connected 25 in turn to a voice activity detector (VAD) 102. The VAD 102 is directly connected to the transmitter 56 (or a controlling processor) with a line 101 and is also interconnected in series with a speech pause/edit unit 104, a speech coder 106, a buffer 108, an optional frame erasure concealment (FEC) unit 110 and a 30 protocol handler 112. In addition to these interconnections, the buffer 108 is also connected to produce a signal 103 back to the VAD 102 and the speech pause/edit unit 104 while the protocol handler 112 produces its output externally to the

11459ROUS01U

transmitter 56.

Figure 3 also shows a typical embodiment for the speech decoder 57. In this embodiment, the speech decoder 57 has connected to the demodulator 65 a protocol handler 114 which 5 is interconnected in series with a jitter buffer 116, an optional FEC unit 110, a speech decoder 120 and a play buffer 122. The play buffer 122 is in turn externally connected to the DAC 66.

This particular embodiment is merely an example 10 illustrating how the speech decoder 57 can be implemented to support downstream communications with the base station 34 (see Figure 1). It is to be understood that other implementations are possible. However, the implementation shown in Figure 3 or 15 any other implementation need not be described here in any further detail as the particular manner in which the speech decoder 57 functions is not material for an understanding of the present invention.

Considering again the speech encoder 55, when 20 digitized speech data is produced by the ADC 52, the digitized data is detected by the VAD 102 which as a result, produces a VAD signal on line 101 denoting the presence of a speech segment. Based on this VAD signal, the transmitter 56 sends an allocation request to the base station 34 to obtain MAC access. The digitized speech data detected by the VAD 102 is immediately 25 forwarded to the speech pause/edit unit 104. After the allocation request has been sent and before a response is received from the base station 34, the speech pause/edit unit 104 proceeds to time compress the speech data received by for example removing repetitive portions present therein and 30 shortening pauses. The time compressed data is then forwarded to the speech coder 106 where it is coded in frames to reduce the speech bit rate (e.g. 64 kbs) to a much lower rate for radio transmission such as for example enhanced full rate codec (ERFC)

11459ROUS01U

at 8kbs. The frames are then stored temporarily in the buffer 108 until transmission.

When a response allocating transmission resources to the mobile station 26 is received from the base station 34, the 5 DSP unit 54 begins to empty the buffer 108 and transmit the frames stored therein. More specifically, after a resource allocation, the frames are forwarded in sequence through the FEC unit 110 to protect against corruption. The frames are then forwarded to the protocol handler 112 where they are placed in 10 packets with one or more frames placed in each packet. The packets are each assembled with an appropriate packet header and sent to the transmitter 56 for transmission to the base station 34.

According to the invention, the transmitter 56 remains 15 operative to transmit speech packets until the VAD 102 detects the end of the speech segment. When this occurs, the VAD 102 sends another VAD signal via line 101 denoting the end of the speech segment to initiate the release of the transmission resources allocated. In the preferred embodiment, the VAD 102 20 is designed with a high sensitivity threshold such that it does not detect any short pauses or periods of silence between speech syllables. A high sensitivity level will reduce the risk that the VAD 102 mistakenly signals the end of a speech segment which has not completed yet.

25 However, the resources will only be released when the segment has completed transmission. More specifically, after detecting the end of a speech segment, the VAD 102 will only initiate a release after being notified via line 103 that the buffer 108 is empty and that the speech segment has completed 30 transmission. In the preferred embodiment, the VAD 102 does not initiate releases immediately as the buffer 108 becomes empty but after a hold time period elapses.

According to the invention, the speech pause/edit unit

11459ROUS01U

104 may be operated to compress the speech data received more or less aggressively, depending on how much time must be saved. According to the preferred embodiment, the speech pause/edit unit 104 should be operated sufficiently aggressively to prevent 5 the buffer 108 from overflowing and therefore losing speech information. In the preferred embodiment, the speech pause/edit unit 104 can monitor the state of the buffer 108 via line 103 and adapt its compressing operations accordingly so that the buffer 108 does not overflow.

10 Further, the speech pause/edit unit 104 should also be operated to provide sufficient time savings to compensate for the additional time required to buffer and transmit frames which would otherwise be discarded if no buffering was used. It can be shown that the additional time required to transmit these 15 frames is equal to the time necessary at the mobile station 26 to acquire resources for transmission. For each speech segment, the speech pause/edit unit 104 should therefore be operated at least long enough to compensate for the resource acquisition time at the mobile station 26. To further reduce transmission 20 delays, the speech pause/edit unit 104 should also be operated long enough to compensate for voice detection delays in the VAD 102.

In other words, the time compression should only be deactivated when the time saved by the speech pause/edit unit 25 104 is equal or greater than the VAD detection time and the time necessary for the mobile station 26 to acquire MAC access. This can also be expressed in the form of an equation as follows:

$$T_{\text{saved}} \geq T_{\text{vad}} + T_{\text{acq}}$$

30

where T_{saved} is the total time saved by the speech pause/edit unit 104, T_{vad} is the speech detection time of the VAD 102 and T_{acq} is the time necessary for the mobile station 26 to

11459ROUS01U

acquire MAC access for a transmission.

To further illustrate this, reference is made to Figure 4 where a timing diagram shows as an example the detection, editing and buffering of a speech segment in the 5 speech encoder 55 of Figure 3 prior to and after its transmission. For clarity, processing delays have been omitted.

The diagram shows labelled as "speech input" the speech segment received by the speech encoder 55. The diagram then illustrates labelled as "VAD" the speech detection in the 10 VAD 102 which occurs after a speech detection time t_{vad} . Next, the diagram shows labelled as "MAC access" the subsequent MAC access by the mobile station 26 after a MAC access time t_{acq} . For the purpose of comparison, the diagram also shows labelled as "onset clipping" the speech segment clipped for a period T_{clip} 15 to illustrate the onset clipping that would occur before MAC access if the speech input was transmitted according to conventional methods.

The diagram then shows labelled as "edited speech" the speech segment compressed by the speech pause/edit unit 104. 20 Next, the diagram shows labelled as "coded and buffered speech" the speech segment coded and buffered prior to its transmission. Finally, the diagram shows labelled as "transmitted speech" the speech segment as transmitted by the transmitter 56.

Before any editing, coding or buffering, the speech 25 segment (see the speech input) is formed of three active speech portions S_1 , S_2 , and S_3 separated by short pauses P_1 and P_2 . At time t_0 , the segment is received in the speech encoder 55. Shortly thereafter, at time t_1 , the VAD 102 detects the segment, generates a VAD signal to initiate a resource allocation request 30 and passes through the speech data detected to the speech pause/edit unit 104 for time compression.

After the VAD 102 detects the speech segment (at time t_1) and for a specified period thereafter, the speech pause/edit

11459ROUS01U

unit 104 compresses the speech data by removing repetitions in the speech portions S_1 , S_2 and S_3 and reducing the pauses P_1 , P_2 (see edited speech). The time compressed speech segment is coded in frames respectively numbered 1 to 11 and stored 5 temporarily in the buffer 108 (see the coded and buffered speech) until MAC access is acquired and the transmitter 56 can begin transmission.

Transmission of the frames begins at time t_2 when MAC access is obtained. At this particular time, the frames 10 contained in the buffer 108 are forwarded in sequence to the transmitter 56 where they are placed in packets for transmission to the base station 34.

From this Figure, it can be observed that by compressing the speech data for a sufficiently long enough 15 period, the speech encoder 55 can catch up on the delays t_{vad} and t_{acq} introduced by speech detection and MAC access (see transmitted speech). By comparison to conventional transmission methods, speech segments transmitted in accordance with the present invention such as shown in Figure 4 can be transmitted 20 without inducing any onset clipping while reducing transmission delays.

The foregoing has described a particular method and apparatus for discarding perceptually insignificant portions in a speech segment and buffering the edited speech to prevent 25 clipping and reduce transmission delays. According to the invention, other techniques can be used.

According to a second embodiment of the invention, the speech data received is first coded in frames and then processed by a frame editor which examines the speech frames and discards 30 frames deemed redundant. The frames which are not discarded by the speech frame editor are placed in a buffer until the mobile station 26 acquires MAC access and begins to transmit.

Figure 5 shows another detailed diagram of the DSP

11459ROUS01U

unit 54 of Figure 2 showing in particular another speech encoder 132 according to the second embodiment of the invention. Similar to the speech encoder 55 of the first embodiment, the speech encoder 132 of this embodiment also has an optional noise reduction unit 140, a VAD 150, a buffer 148, an optional FEC unit 150 and a protocol handler 152. These elements are interconnected between the ADC 52 and the transmitter 56 in the same manner the optional noise reduction unit 100, the VAD 102, the buffer 108, the FEC unit 110 and the protocol handler 112 of the speech encoder 55 are interconnected. In contrast to the speech encoder 55 however, the speech encoder 132 has between the VAD 150 and the buffer 148 a speech coder 144 interconnected in series with a speech frame editor 146.

In this particular embodiment, when a user initiates a speech segment, the digitized speech data produced by the ADC 52 and passed through the optional noise reduction unit 140 is detected by the VAD 142. Similarly to the VAD 102, the VAD 142 is also designed with a high sensitivity threshold to remove the risk of mistakenly signaling the end of speech segments.

Upon detecting the digitized speech, the VAD 142 sends a VAD signal directly to the transmitter 56 via a line 141 denoting the presence of a speech segment. Based on this VAD signal, the transmitter 56 sends a resource allocation request to the base station 34 to obtain MAC access. The digitized speech data detected by the VAD 102 is sent to the speech coder 144 to be coded into frames and the frames are then forwarded to the speech frame editor 146 for editing.

After the resource allocation request has been sent and before a response is received from the base station 34, the speech frame editor 146 proceeds to remove perceptually insignificant portions in the coded speech by discarding frames which it deems redundant. This could be for example, frames which contain repetitive speech portions or short pauses. The

11459ROUS01U

frames which are not discarded by the speech frame editor 146 are then placed in the buffer 148 until the mobile station 26 obtains a resource allocation and can begin transmission of the segment.

5 When a response allocating transmission resources to the mobile station 26 is received, the frames stored in the buffer 148 are passed through the FEC unit 150 to protect against corruption. After passing through the FEC unit 150, the frames are then forwarded to the protocol handler 152 in sequence where they are placed in packets and sent to the transmitter 56 for transmission to the base station 34.

In this particular embodiment, the speech frame editor 146 should also be operated to provide sufficient time savings to compensate for the speech detection and resource allocation 15 delays. More specifically, the speech frame editor 146 should only be disabled when the time saved T_{saved} is equal or greater than the VAD detection time t_{vad} and the MAC access acquisition time t_{acq} (see above equation for T_{saved}).

This is further illustrated in Figure 6 in which a 20 timing diagram shows as an example the speech segment of Figure 5 as detected, frame edited and buffered by the speech encoder 132 prior to and after its transmission (processing delays omitted). For clarity, this diagram reproduces the speech segment as received by the speech encoder 132 together with the 25 VAD timing of the VAD 142 and the MAC access by the mobile station 26 following the detection of the speech input. In addition, and for the purpose of comparison, the diagram shows labelled as "onset clipping" the speech segment clipped for a period T_{clip} to illustrate the onset clipping that would occur if 30 the speech input in this particular example was transmitted according to conventional methods.

The diagram then shows labelled as "coded speech" the speech segment coded in frames by the speech coder 144. Next,

11459ROUS01U

the diagram shows labelled as "edited and buffered speech" the speech segment edited and buffered prior to its transmission. Finally, the diagram illustrates labelled as "transmitted speech" the speech segment as transmitted by the transmitter 56.

5 As in Figure 4, the speech segment shown here has three active speech portions S₁, S₂, and S₃ separated by pauses P₁ and P₂. Again, at time t₀, the segment is received in the speech encoder 132. At time t₁, the VAD 142 detects the segment, generates a VAD signal for initiating the allocation request to
10 the base station 34 and passes the speech data detected to the speech coder 144. There, the speech is coded in frames. In this particular example, the speech segment is coded in eleven frames respectively numbered 1 to 13.

The frames generated by the speech coder 144 are
15 passed through the speech frame editor 146 which is operative to discard redundant frames in the segment which may contain for example repetitive speech portions or short pauses. In the example shown, frames 6 and 10 are discarded because they contain short pauses. The frames which are not discarded by the
20 speech frame editor 146 are placed in the buffer 148 until the mobile station 26 obtains a resource allocation and can begin transmission of the segment. Again, transmission of the frames begins at time t₂ when MAC access is obtained. At this particular time, the frames contained in the buffer 148 are
25 forwarded in sequence to the transmitter 56 where they are placed in packets for transmission to the base station 34.

From this Figure, it can be observed that by discarding redundant frames for a sufficiently long period and providing adequate buffering, the speech encoder 132 can also
30 catch up on the speech detection and resource allocation delays t_{vad}, t_{acq} and transmit segments without inducing any onset clipping.

Figure 7 illustrates a sample protocol stack that may

11459ROUS01U

be used in the network 10 of Figure 1 to transmit speech segments according to the preferred embodiments of the invention described above. The protocol stack shown in this figure consists of an application layer, a transport protocol layer, a 5 network protocol layer, a link layer and a physical layer. In the well-known open systems interconnection reference (OSI) model, these layers are respectively referred to as layers 7, 4, 3, 2 and 1.

In this particular stack example, the speech detection 10 and coding performed by the speech encoder 55, 132 is implemented in the application layer level (layer 7). In the transport protocol layer (layer 4), the network 10 uses a real-time transport protocol (RTP) and a user datagram protocol (UDP). The RTP protocol is a packet format protocol used in the 15 network 10 to transmit multimedia streams. This particular protocol utilizes existing transport layers for data such as voice which has real-time properties and time constraints. The network 10 also uses a UDP protocol in the transport layer which runs below RTP. UDP is a transport layer protocol which 20 functions as a best effort protocol without guarantee of delivery.

In the network protocol layer (layer 3), the network 10 uses an Internet protocol (IP) for routing user data and control signalling. In the link layer (layer 2), the network 10 25 uses a subnetwork dependent convergence protocol (SNDCP) to allow transfer of higher network layer protocol data units in a transparent manner. The SNDCP protocol performs multiplexing of these data units for transmission using the service provided by the logical link control (LLC) protocol. The LLC protocol 30 conveys information between layer 3 entities in the mobile stations 26, 28 and elsewhere in the network 10. Below the LLC protocol, the network 10 uses a radio link control (RLC) protocol which defines the procedure for segmentation and re-

11459ROUS01U

assembly of layer 2 packet data units. Also used in the link layer is a MAC protocol which multiplexes users onto a shared transmission medium. In each cell 12, 14, 16, 18, 20, 22, 24 of the network 10, this shared transmission medium is consists of 5 transmission resources. In the physical layer (below the MAC protocol), the network 10 has defined all the physical elements used communications in each cell 12, 14, 16, 18, 20, 22 and 24 between mobile stations 26, 28 and base stations 32, 34, 36, 38, 40, 42, 44. This includes transmitters, receivers and the 10 transmission resources used in each cell 12, 14, 16, 18, 20, 22 and 24.

Generally, network devices and elements located in a particular layer only exchange messages with devices or elements located in the same or an adjacent layer. As is conventional, 15 the preferred embodiment of the invention introduces an exception to this with respect to the speech detection by the VAD 102 (or the VAD 142) and more particularly to the generation of a VAD signal by the VAD 102. It will be recalled that the VAD 102 produces a VAD signal 101 to the transmitter 56 to 20 initiate resource allocation requests when speech is detected. Because the VAD 102 must produce this VAD signal directly to the transmitter 56, the application layer in this protocol stack example is shown as being capable of communicating directly to the physical layer.

25 While the invention has been described with reference to a particular multi-access system, further modifications and improvements to apply the invention in other types of multi-access systems which will occur to those skilled in the art, may be made within the purview of the appended claims, without 30 departing from the scope of the invention in its broader aspect.

Further, the invention has been described above in relation to the upstream transmission of voice from a mobile station to a base station. It is to be understood that the

11459ROUS01U

invention could also be used in the downstream transmission of voice from the base station to the mobile station.

More generally, the invention can be used in different multi-access systems for a variety of audio services such as for example, Enhanced Data for Global Evolution and voice over Internet Protocol (VoIP) services. In addition, the invention can also be used for other real-time services such as for example video conferencing services. However, should the invention be used in the transmission of information other than speech, it becomes apparent that there may be elements among those described above which may have to be reconfigured or replaced by components suited to handle the type of information sought to be transmitted.

For example, the VAD 102, the speech pause/edit unit 15 104 and the speech coder 106 of Figure 3 are components which are used in relation to the transmission of speech. For other types of information (other than speech), different components performing the same functions but adapted to the particular type of information to transmit would have to be used. It is to be 20 understood that these components could be described more generally as an information detector, an information editor and a coder respectively. Similarly, the VAD 142, speech coder 144 and the speech frame editor 146 of Figure 5 could also be more generally described as an information activity detector, a coder 25 and an information editor.

Also, the invention is not restricted to the particular protocol stack example described above. It is to be understood that other protocol stacks could be used. A different protocol stack could be used where for example, 30 different protocols are used for communications in the network 10.

The invention has been described above in relation to a particular resource allocation scheme whereby transmission

11459ROUS01U

resources are only allocated for active speech segments. It is to be understood that the invention can also be used in multi-access systems with other types of resource allocation mechanisms. For example, the invention can also be used in 5 systems using a pre-emptive mechanism where transmission resources are allocated upon request but with no need for a response.

According to the invention, the steps described above in relation to the processing of each speech segment in the 10 speech encoder are not to be understood to be strictly applied in the order described above. For example, with respect to the steps of editing and buffering, each speech segment can be edited first and then buffered before they are transmitted. Conversely, the segments could each be buffered first and then 15 edited prior to transmission. Further, it is to be understood that processing steps may be performed on different versions of the segments and still fall within the purview of the invention. For example, the editing and buffering steps may be performed before or after the segments are coded in frames.